

2008

Efficiency gains for seasonal adjustment by joint modelling of disaggregated series

Carole Birrell

University of Wollongong, cbirrell@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/theses>

University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Recommended Citation

Birrell, Carole, Efficiency gains for seasonal adjustment by joint modelling of disaggregated series, PhD thesis, School of Mathematics and Applied Statistics, University of Wollongong, 2008.
<http://ro.uow.edu.au/theses/131>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Efficiency Gains for Seasonal Adjustment by Joint Modelling of Disaggregated Series

*A thesis submitted in fulfillment of the
requirements for the award of the degree*

Doctor of Philosophy

from

University of Wollongong

by

Carole Birrell BSc (UNSW), DipEd, MStat (Wollongong)

School of Mathematics and Applied Statistics

2008

CERTIFICATION

I, Carole Louise Birrell, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Mathematics and Applied Statistics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Carole Louise Birrell

3 August, 2008

Abstract

Governments and businesses use data collected over time as indicators of the social, economic and business conditions of the country. These may then be used for policy and planning decisions, calculation of national accounts and monitoring of economic activity. The production and publication of seasonally adjusted series, in addition to unadjusted figures, is standard practice for government statistical agencies. In general, there are two main approaches to seasonal adjustment, namely a filter-based approach and a model-based approach. Filter-based methods estimate time series components, such as the trend and seasonal factors, by application of a set of filters to the original series. Model-based methods of seasonal adjustment are more specific to each series, and are thereby more flexible.

Time series resulting from aggregation of several sub-series can be seasonally adjusted directly or indirectly. With model-based seasonal adjustment, the sub-series may also be considered as a multivariate system of series and hence the analysis may be done jointly. This approach has considerable advantage over the indirect method, as it utilises the covariance structure between the sub-series.

The focus of this thesis is on examining how the accuracy of seasonally adjusted series can be improved by using the sub-series. A model-based approach to seasonally adjusting an aggregated series is carried out with two different methods. The first method utilises an univariate basic structural model (BSM) for the aggregated series. The second method utilises a multivariate basic structural model for the sub-series. In basic structural models, the series components are modelled individually, and then put into state space form. The Kalman filter is applied to obtain estimates of the aggregate series components and the prediction mean squared errors.

The variance of the seasonally adjusted series given by the two methods is studied through their relative efficiency. A particular emphasis of the thesis is on how the similarity of and differences between disaggregated series affect the efficiency of the two approaches to seasonal adjustment.

Results indicate that gains are attainable under specified conditions which rely on the values of the parameters of not only the seasonal component, but also the non-seasonal components. These results demonstrate the impact on relative efficiency of relationships among sub-series parameters, both between series (i.e. within components) and within series (i.e. between components).

The impact of the length of the time series on the accuracy of seasonally adjusted series is of particular interest. A simulation study investigates the parameter estimates obtained given varying series lengths and the subsequent effects on the accuracy of the time series components given by the Kalman filter. These effects are measured by the naïve bias in the prediction mean squared error and by the revision error. A bootstrap correction is applied to the estimated prediction mean squared error for both the univariate and multivariate approaches.

A single indicator measure is developed for predicting whether the properties of the disaggregated series (or sub-series) will lead to gains in the accuracy of the seasonally adjusted aggregated series. The quasi-likelihood method is applied to obtain the indicator measure of relative efficiency. It is shown to be directly related to the relative efficiency measure obtained with the Kalman filter.

Another application of the quasi-likelihood indicator is in identifying an appropriate grouping of the K sub-series into $r < K$ series. The grouping can considerably reduce the number of estimated parameters, while the accuracy of the seasonally adjusted series is maintained.

The integrated approach of this thesis to the seasonal adjustment of aggregated series thus provides a pathway to improved efficiency and an understanding of the conditions under which improvements may be achieved.

Acknowledgements

This thesis would not have been possible without the help and support of many people. I would firstly like to thank my supervisors, Professor David G. Steel and Associate Professor Yan-Xia Lin. Thank you David for your wisdom, knowledge, guidance and great patience. I appreciate that your comments have always been positive and for believing that I would finish this thesis one day, even when I had trouble believing it would happen. Thank you also for the opportunities to travel to Berlin, Southampton and Luxembourg. To Yan-Xia, thank you for the time you have spent working through most of the detail of this thesis. Your help with the theory has been invaluable and I have appreciated all the discussions we have had to get to this point. You have taught me so much.

This project has been funded by the Australian Research Council and the Australian Bureau of Statistics. I would especially like to thank Dr Mark Zhang and Dr Craig McLaren of the Time Series Analysis division for their advice early on in this project and for the time they made available to meet with me on my visits to Canberra.

On my visit to Southampton, I had the opportunity to visit academics including Professor Danny Pfeffermann. Thank you for your suggestions and help with the direction of this thesis.

The journey of a PhD is more enjoyable if it can be shared with others in the same situation. Thanks to all my fellow PhD students, particularly Sandy, Elahe, Maureen and Raed, for sharing the experiences. Thanks especially to Sandy for your friendship and all the coffees we had together when we needed an excuse to have a break. Your advice and help have been much appreciated.

I would also like to thank the staff in the Department of Mathematics and Applied Statistics for their support and encouraging words over many years. In particular, a special thank you to Professor David Griffiths who encouraged me to come back to study after a long break. Your friendship, support and encouragement is greatly valued. Your willingness and time taken to proof-read the thesis is very much appreciated.

To my friends at Figtree Anglican Church and Bible study, thanks for your fellowship and prayers, and for understanding when things were so busy. Your loyalty and support is what makes you such great friends. Special thanks to Alison and the Kerr family for the times we have shared and for the many retail therapy sessions! Thank you to Kristina, for although our lives have taken different paths over these last few years, our friendship has only grown.

To my parents and family, thanks for your understanding, for always believing in me and spurring me on to get to the finish line. I know I wouldn't have achieved this without your support.

Most importantly, to my husband Michael, for the unfailing love, patience, support and encouragement you have always given me. Thank you for your graciousness in helping with the cooking, cleaning, the occasional programming, and for setting up the computers at home. I greatly appreciate all the little things you have done, such as making me coffee and bringing me chocolate to keep me at my desk! Thanks for the talks, for the walks on the beach and for just being there at the end of every day.

I thank my Lord Jesus Christ for sustaining me throughout this time. In the words of Ecclesiastes 3:1, "There is a time for everything and a season for every activity under heaven" (Holy Bible, New International Version). I am looking forward to the next season in my life with Him at my side. I know that whatever it may be, "I can do everything through Him who gives me strength" (Holy Bible, New International Version, Philippians 4:13).

Contents

1	Introduction	1
2	Issues in Seasonal Adjustment	7
2.1	Current Methods	8
2.1.1	Filter-based Seasonal Adjustment: X-11	8
2.1.2	Model-based Seasonal Adjustment	10
2.1.3	Parameter Estimation and Bias	13
2.2	Comparison of X-11 with Model-Based Estimates	15
2.3	Measuring Revision Error	17
2.4	Seasonal Adjustment of an Aggregate Series	18
2.5	Applications of the Multivariate Basic Structural Model	23
2.6	Short Time Series	24
2.6.1	Short Time Series and Longitudinal Analysis	26
2.7	Modelling Survey Errors within a State Space Model Context	27
2.8	Intervention Analysis	28
2.9	Conclusion	30
3	Modelling Disaggregated Series	32
3.1	Introduction	32
3.2	Univariate BSM	33
3.2.1	The Trend Component	33
3.2.2	The Seasonal Component	34
3.2.3	The Disturbance Terms	35

3.3	Multivariate BSM	35
3.3.1	Placing Restrictions on the Variance Structure	37
3.4	State Space Form	42
3.4.1	Univariate BSM in State Space Form	42
3.4.2	Multivariate BSM in State Space Form	44
3.4.3	Applying a Transformation	46
3.5	Application of the Kalman Filter	49
3.5.1	Kalman Filter for the Univariate Model	49
3.5.2	Kalman Filter for the Transformed Multivariate Model	51
3.5.3	Kalman Smoother	52
3.6	Comparison of Univariate and Multivariate Methods	54
3.6.1	Variance of the Seasonally Adjusted Series	54
4	Empirical Study with Exact Parameters	57
4.1	Introduction	57
4.2	Model Structure	58
4.2.1	State Space Form	60
4.3	Design of the Study	62
4.3.1	Setting the Parameters	63
4.3.2	Application of Constraints	66
4.4	Estimation of Components via the Kalman Filter	67
4.5	Results	68
4.5.1	The Effect of the C-ratios	68
4.5.2	The Effect of Correlation Settings	73
4.6	Other Aggregated Series	81
4.6.1	Results for other Aggregated Series	83
4.7	Conclusion	88
5	Quasi-Likelihood Indicator of Relative Efficiency	92
5.1	Introduction	92
5.2	Introduction to Quasi-Likelihood Method	94

5.3	Quasi-Likelihood Approach to the Univariate BSM	95
5.4	Quasi-Likelihood Approach to the Multivariate BSM	100
5.4.1	Transforming the Multivariate System	102
5.5	The Quasi-Likelihood Indicator	109
5.6	Comparison with Kalman Filter Results	110
5.6.1	Homogeneous Model	111
5.6.2	Model 1: Compound Symmetry	114
5.6.3	Model 2: Unit-specific Variances	116
5.7	Grouping Application	121
6	Parameter Estimation: Varying Series Length and Accuracy of Es- timated Components	127
6.1	Introduction	127
6.2	Estimation of Parameters in the State Space Model	128
6.2.1	Univariate Model	130
6.2.2	Multivariate Model	133
6.3	Simulation Experiment	139
6.3.1	Generation of Data Series	139
6.3.2	Estimation of Parameters	142
6.4	Results of Parameter Estimates for Varying Series Length	144
6.4.1	Design <i>A1a14</i>	145
6.4.2	Design <i>A1b33</i>	151
6.5	Effect of Parameter Estimation on MSE of Model Components	157
6.5.1	Naïve Bias	159
6.5.2	Bias Correction	164
6.6	Revision Error	170
6.6.1	Revision Error Results	172
6.7	Summary	175
7	Handling Several Sub-series: An Example	177
7.1	Introduction	177

7.2	Data Generation of the Eight Sub-series	178
7.3	Grouping the Sub-series	184
7.3.1	Calculation of Q for each Combination	187
7.4	Relative Efficiency Results	189
7.5	Summary	191
8	Conclusions	193
8.1	Summary	193
8.2	Future Directions	196
A	The Kronecker Product	199
B	Trigonometric Seasonality	201
B.1	Introduction	201
B.2	Univariate BSM	201
B.2.1	Univariate State Space Form	203
B.3	Multivariate BSM	204
B.3.1	Multivariate State Space Form	205
B.4	Calculation of Relative Efficiency	206
B.5	Relative Efficiency Results	207
C	Derivations for Chapter 5	208
C.1	Quasi-Likelihood Result for the Error Variance Matrix of the Seasonal Components	208
C.2	Denominator of the Quasi-Likelihood Indicator of Relative Efficiency	212
C.3	Stirling Numbers of the Second Kind	215
D	Derivations for Chapter 6	217
D.1	Diffuse Log-likelihood	217
D.1.1	Exact Initial Kalman Filter	217
D.2	Derivations for Section 6.2.1	219

D.2.1	Calculation of w_t	219
D.2.2	Concentrated Diffuse Log-likelihood	220
D.3	Derivations for Section 6.2.2	220
D.3.1	Calculation of w_t	220
D.3.2	Concentrated Diffuse Log-likelihood	222
E	Program Files	223
E.1	Chapter 4 Programs	224
E.2	Chapter 5 Programs	230
E.3	Chapter 6 Programs	231
E.3.1	Programs for Section 6.4.1: Design <i>A1a14</i>	231
E.3.2	Programs for Section 6.4.2: Design <i>A1b33</i>	233
E.3.3	Programs for Section 6.5.1: Naïve Bias	235
E.3.4	Programs for Section 6.5.2: Bootstrap Correction	236
E.3.5	Programs for Section 6.6: Revision Error	237
E.4	Chapter 7 Programs	238
E.5	Appendix B Programs	242
E.6	Appendix D Programs	242
	Bibliography	243

List of Figures

4.1	$RE_t(M)$ for sub-series design $A1a$.	70
4.2	$RE_t(M)$ for sub-series design $A1b$.	71
4.3	$RE_t(M)$ for design $a12$ with $A1 - A5$.	73
4.4	$RE_t(M)$ for design $a13$ with $A1 - A5$.	74
4.5	$RE_t(M)$ for design $a14$ with $A1 - A5$.	74
4.6	Seasonal correlation versus $RE_{40}(M)$ for designs $a11, a12, a13, a14$ with $A1-A5$.	75
4.7	$RE_t(M)$ for design $a21$ with $A1 - E1$.	76
4.8	$RE_t(M)$ for design $a31$ with $A1 - E1$.	77
4.9	$RE_t(M)$ for design $a41$ with $A1 - E1$.	77
4.10	Non-seasonal correlation versus $RE_{40}(M)$ for designs $a11, a21, a31,$ $a41$ with $A1 - E1$.	78
4.11	$RE_t(M)$ for designs $A1a11$ to $A1a44$ given by <i>Series 1, 2</i> and <i>3</i> .	86
4.12	$RE_t(M)$ for designs $A1b22$ to $A1b44$ given by <i>Series 1, 2</i> and <i>3</i> .	87
5.1	Quasi-likelihood indicator vs KF relative efficiency	120
6.1	Box plot of the level parameter estimate for design $A1a14$.	146
6.2	Box plot of the seasonal parameter estimate for design $A1a14$.	147
6.3	Box plot of the measurement error parameter estimate for design $A1a14$.	147
6.4	Box plot of the level parameter estimate for design $A1b33$.	151
6.5	Box plot of the seasonal parameter estimate for design $A1b33$.	152
6.6	Box plot of the measurement error parameter estimate for design $A1b33$.	152
6.7	Box plot for relative efficiency of revision error variance	174

6.8	Box plot for relative efficiency of revision error variance with limited scale	174
7.1	Logarithms of quarterly wine sales by type	180
7.2	Simulated series of the logarithms of quarterly wine sales by type . .	183

List of Tables

4.1	Labels for sub-series design ‘ <i>a</i> ’.	64
4.2	Labels for sub-series design ‘ <i>b</i> ’.	65
4.3	Labels for correlation combinations.	65
4.4	Results of $RE_{40}(M)$ for sub-series design <i>A1a</i> .	70
4.5	Results of $RE_{40}(M)$ for sub-series design <i>A1b</i> .	71
4.6	Results of $RE_{40}(M)$ for design ‘ <i>a</i> ’ correlation combinations.	79
4.7	Results of $RE_{40}(M)$ for design ‘ <i>b</i> ’ correlation combinations.	81
4.8	<i>Series 1, 2, and 3</i> : results of $RE_{40}(M)$ for design <i>A1a</i> .	85
4.9	<i>Series 1, 2, and 3</i> : results of $RE_{40}(M)$ for design <i>A1b</i> .	85
5.1	Results of $RE_{40}(M)$ and Q for sub-series design ‘ <i>a</i> ’: part 1.	117
5.2	Results of $RE_{40}(M)$ and Q for sub-series design ‘ <i>a</i> ’: part 2.	119
5.3	Results of $RE_{40}(M)$ and Q for sub-series design ‘ <i>b</i> ’.	120
6.1	Parameters for sub-series design <i>A1a14</i> and <i>A1b33</i> .	140
6.2	Median parameter estimates for <i>Series 2</i> for design <i>A1a14</i> .	148
6.3	Means and standard errors of parameter estimates for design <i>A1a14</i> .	149
6.4	Relative efficiency of the variances of the parameter estimates for design <i>A1a14</i>	150
6.5	Median parameter estimates for design <i>A1b33</i> .	154
6.6	Mean and standard errors of parameter estimates for design <i>A1b33</i> .	155
6.7	Relative efficiency of the parameter estimates for design <i>A1b33</i> .	156
6.8	Mean and RMSE of relative naïve bias for design <i>A1a14</i> .	163
6.9	Mean and RMSE of relative naïve bias for design <i>A1b33</i> .	163

6.10	Naïve bias results for design $A1b33$ with $T = 28$	169
6.11	Bootstrap correction for design $A1b33$ with $T = 28$	169
6.12	Relative efficiency of the revision error variance for design $A1a14$ and $A1b33$	173
7.1	Parameters used for generation of the eight sub-series	181
7.2	Initial values used for generation of the eight sub-series	182
7.3	C-ratios for pairs of series	186
7.4	Different combinations of the eight sub-series	187
7.5	Values of Q and relative efficiency for nine different combinations . .	189
B.1	Results for $RE_{40}(M)$ for trigonometric seasonal component.	207
C.1	Stirling numbers of the second kind	216

List of Abbreviations

ABS	Australian Bureau of Statistics
AMB	ARIMA-model-based
AR	autoregressive
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
BLS	Bureau of Labor Statistics
BSM	basic structural model
CPS	Current Population Survey
IGLS	iterative generalised least squares
KF	Kalman filter
KS	Kalman smoother
LLS	local level seasonal
MSE	mean squared error
QL	quasi-likelihood
PMSE	prediction mean squared error
RMSE	root mean squared error
SEA	survey error autocorrelations
SEATS	Signal Extraction in ARIMA Time Series
SSF	state space form
STAMP	Structural Time series Analyser, Modeller and Predictor
SUTSE	seemingly unrelated time series equations
TRAMO	time series regression with ARIMA noise, missing values, and outliers
UC	unobserved components